
STATISTICAL INFERENCE USED IN MARKETING RESEARCH

CĂTOIU Iacob

Academy of Economic Studies, Bucharest, Romania

STANCIU Oana

Lucian Blaga University of Sibiu, Romania

ȚICHINDELEAN Mihai

Lucian Blaga University of Sibiu, Romania

Abstract:

The purpose of this article is to describe, theoretically, but also through several applications, the statistical methods or procedures through which the research results can be extended at the target population level. The target population represents the aggregate of all the elements, sharing some common set of characteristics that comprise the universe for the purpose of the marketing research problem

Keywords: *marketing research, sample, statistical inference, statistic test*

1. Introduction

The purpose of every human being is to be happy during his lifetime. Companies are constituted out of people, but their happiness purpose relies in profit maximization and obtaining customer satisfaction. These two variables are directly linked one to another, obtaining customer satisfaction is the cause (and not the effect) for companies' profit. Marketing – as theory and practice – tries to find a balance point, equilibrium between these two company's goals, which apparently form a concave relation which needs empirical confirmation.

To measure the degrees of customer satisfaction, the company needs relevant information about their customers. The needed information is based on data about the customers' initial expectations, regarding the company's performance, and the final customer's perception about the company's performance. Measuring and analyzing data is part of the marketing research process. Marketing research is defined as a key element within the total field of marketing information. It links the consumer, customer

and the public to the marketer through information, which is used to identify and define marketing opportunities and problems; to generate, refine and evaluate marketing actions; and to improve the understanding of marketing, as a process and of the way in which specific marketing activities can be made more effective (Malhotra N., Birks D., 2007). The scope of every marketing research is to deliver information, which may assist or help the decider in his/her decision. Thus, marketing research is viewed as a process, a sequence of phases and activities, which contribute to the achievement of the research's purpose and to the solution of a certain marketing decision problem (Catoiu, 2009).

According to the research purpose, the researcher will collect the needed data by using one or more of the marketing data gathering methods. There are four primary research methods through which original data (related to the actual research's purpose) is gathered: the survey (method based upon the use of structured questionnaires administered to a sample population - Malhotra); the interview and the focus group (qualitative methods by which the researcher attracts qualitative data through discussion with one or more investigated subjects) and the observation (the recording and counting of behavioral patterns of people, objects and events in a systematic manner in order to obtain information about the phenomenon of interest - Malhotra). Other data gathering methods use secondary data (existing data) as a starting point for the future analysis: the simulation (a bundle of research methods and techniques of a real system S with the use of a simulated system S', with which it has an analogue relation - Catoiu), the experiment (research method which controls and manipulates one variable's variation – independent variable – with the purpose of measuring its effects on a second variable's variation – dependent variable - Catoiu) and the analysis of secondary data.

Each data gathering method uses information (primary and/or secondary data) from an investigated subject or data source. The number of investigated subjects does rarely equal the targeted population (there are rare cases in which the researcher undertakes a census); therefore a sample is drawn out of the target population. Based on the results of the sample data, the researcher is tempted to extend (to infer) his conclusion at the target population level. This inference should always be made on statistical basis.

2. Research methodology

The purpose of this article is to describe, theoretically, but also through several applications, the statistical methods or procedures through which the research results can be extended at the target population level. The target population represents the aggregate of all the elements, sharing some common set of characteristics that comprise the universe for the purpose of the marketing research problem (Malhotra, 2007). A census is a complete enumeration of the elements of a population or study's

objects, while a sample is defined as a subgroup of elements of the target population. Relevant results based on the sample data are obtained if the researcher rigorously respects all phases of the sampling process: defining target population, determining sample frame, choosing a sampling method, determining a sample size and testing the used sampling methods (Catoiu 2009). Within this article, based on the known sample size (although it takes different values), statistical methods are described and used in order to verify the relevance of the sample results for the target population.

Statistical inference is the process through which the researcher has to deduce or infer facts about the target population from which the sample is drawn (Thomas 1997). The problem in statistical inference relies in the variability of the sampling process, because different samples will yield different results. Therefore, a sample of size “n” is called a randomly drawn sample if every combination of n-subjects of the target population has equal chances to become the actually drawn sample.

If μ is the population mean and σ^2 its variance, and a n-sized sample is drawn out of the population, then through statistical inference $E(X_s) = \mu$ – expectation of the sample mean is the population’s mean – and $\text{var}(X_s) = \sigma^2$ – expectation of the sample variance is the population’s variance. According to the Central Limit Theorem (CLT), if large sized samples are randomly drawn out of the target population with the mean μ and the variance σ^2 , then the sampling distribution’s mean is a normal distribution, regardless of the initial distribution of the population from which the samples are drawn.

Therefore:

Mean $E(X_s) = \mu$ and $\text{var}(X_s) = \sigma^2$

Estimation of the population’s mean

If a parameter related to the population (mean or variance) is unknown, it can be estimated through two methods: Number (point estimate) and interval (confidence interval). In the following the authors present the first method – of point estimate.

Estimation of the population’s mean relies on the known sample mean (X_s). If a very large number of samples would be drawn from the target population, then the distribution of the sample mean will rely on the probability density function of the population, like:

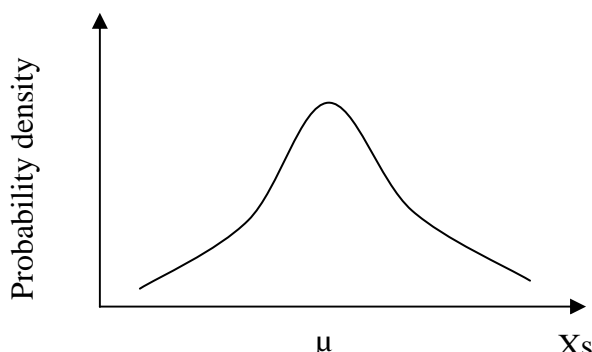


Figure 1 – Normal distribution

The mean value of the sample mean is the population mean: mean $E(X_s) = \mu$. It can be said that the point estimate of the population mean is not affected by systematic error; therefore he is an unbiased estimate.

To exemplify the up mentioned theory, a target population was chosen of 200 companies characterized by their revenues obtained in the year 2010 (http://money.cnn.com/magazines/fortune/fortune500/2010/full_list/301_400.html).

By using the procedure Random Sample of the statistic software PASW 18, a first random sample was drawn with $n=60$ companies. The population mean is $\mu = 38.658\$$, while the sample mean is $X_s = 46.667\$$. (Table)

Table 1. Statistics for a $n=60$ sample. One-Sample Statistics

	n	Mean	Std. Deviation	Std. Error Mean
Revenues	60	46677,22	56867,598	7341,575

The two mean values are different, but how much different. Based on the difference between the two values, two hypotheses are developed which can confirm or not if the mean of the sample drawn differs significantly from the population mean.

Thus, a null hypothesis indicates that there is no significant difference between the sample mean and the population mean:

$$H_0: \mu = 46.677\$$$

While an alternative hypothesis states that there is a significant difference between the sample mean and the population mean:

$$H_1: \mu \neq 46.677\$$$

According to CLT, \bar{X}_s is following a normal distribution with mean μ and variance σ^2/n . To test which of the hypotheses will be accepted, statistical tests are used based on tabular values of the standardized normal distribution (z-values) and of the t-Student distribution (t-values).

Assuming that H_0 is true, namely that the mean value of the sample revenues does not significantly differ from the mean value of the population revenues, a test statistic:

$$T.S. = \frac{\bar{X}_s - \mu}{\frac{\sigma}{\sqrt{n}}} \quad \text{will follow a normal standardized distribution } N(0,1).$$

If H_0 is accepted, then the computed values should lie within the marked interval:

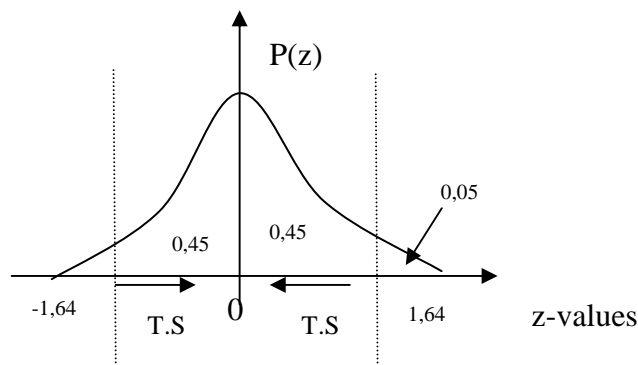


Figure 2 – Standardized normal distribution

The area between the curve and the Ox axis (to the right of a z-value of 1.64 and to the left of a z-value of -1.64) is 0.1, which is noted by α and is called “significance level”. If the value of the T.S. lies between the interval $(-1.64, 1.64)$ we can reserve our judgment regarding H_0 that the sample mean is the same with the population mean. If it does not lie within the mentioned interval, we can reject H_0 (the sample mean differs significantly from the population mean) with a probability of 0.1 to falsely reject H_0 , which means that there is a 1 to 10 chance that the mean of another n-sized sample does not differ from the population mean.

In our example,

$$T.S. = \frac{46.677 - 38658}{\frac{56.867}{\sqrt{60}}} = 1.091$$

Table 2. Statistics of t-Student test for a n=60 sample One-Sample Test

	Test Value = 38658					
	t	Df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
Revenues	1,092	59	,279	8019,217	-4249,24	20287,67

We compare the computed value of the T.S. with a tabular value for the standardized normal distribution (the value is chosen based on the desired level of significance, for $\alpha=0,1$ $z_{0,1}=1.64$) and a tabular value for the t-Student distribution (for $\alpha=0,1$ and $r=n-1=59$ degrees of freedom $t_{0,1,59}=1,296$).

T.S. = 1.091 < $z_{0,1}=1.64$ results that there is insufficient evidence to reject H_0 , but can say with a probability of 0,1 of being wrong that the mean of the sample revenues does not differ significantly from the mean population revenues.

T.S. = 1,091 < $t_{0,1,59}=1.296$ implies not a rejection of H_0 , because of lack of evidence. Table no.2 contains a value of 0.279 (Sig. 2-tailed) which represents the probability of 0.279 of rejecting H_0 when she is true.

If we draw out of the population a random sample $n=5$ companies, we obtain the following results:

Table 3. Statistics for a n=5 sample. One-Sample Statistics

	n	Mean	Std. Deviation	Std. Error Mean
Revenues	5	23642.00	11857.794	5302.967

We want to test again if the sample mean differs significantly from the population mean.

Thus: $H_0: \mu=46.677\$$ and
 $H_1: \mu \neq 46.677\$$

$$T.S. = \frac{23.643 - 38.658}{\frac{11.857}{\sqrt{5}}} = -2.836 \text{ (Table 4)}$$

Table 4. Statistics of t-Student test for a n=5 sample. One-Sample Test

	Test Value = 38658					
	t	df	Sig. (2-tailed)	Mean Difference	90% Confidence Interval of the Difference	
					Lower	Upper
Revenues	-2.832	4	.047	-15016.000	-26321.11	-3710.89

T.S. = $-2.836 < z_{0,1} = -1,64$ indicates that there is enough evidence to reject H_0 , at a significance level of 0.1. The negative sign of the T.S. indicates that the sample mean is below the population mean.

We reached the same conclusion by comparing T.S. = $-2.836 < t_{0,1,4} = -2.132$. The value of 0.047 (Sig. 2-tailed) indicates that there is a probability of 0,047 of rejecting H_0 , when it is true.

3. Conclusion and future research

The main purpose of the paper is to describe, theoretically, the importance of statistical inference as a process used within marketing research. The theoretical background is supported by an example through which the authors demonstrate, empirically, that the sample results cannot always be extrapolated to the entire researched population. The main factor which influences this inference (by using statistic tests) is the sample size. The bigger the sample size is, the more representative the sample results are, the bigger is the probability of rejecting H_0 , when it is true. In testing the power of a population estimate, the researcher should use different statistic tests, based on the sample size, if the sample size is small ($n < 30$), one should use t-Student test, otherwise ($n \geq 30$) the standardized normal distribution test should be followed.

Future research will be orientated towards an empirical conformation of the theory regarding the estimate of the population variance. Furthermore, the authors will compare the strengths and weaknesses of the known sampling methods.

References

- Burns A., Bush R.(2006): Marketing Research, Pretince Hall, New Jersey
 Cătoi I. (coordonator)(2009): Tratat de Cercetări de marketing, Editura Uranus, București
 Greene W., (2002): Econometric Analysis, Pretince Hall, New Jersey
 Kane E.J.:Economic Statistic and Econometrics, Harper and Row, New York
 Malhotra N., Birks D.(2007): Marketing Research, An Applied Approach, Pretince Hall, Essex
 Thomas R.L. (1997): Modern Econometrics, Addison-Wesley, Essex